



Denoising of stimulated Raman scattering microscopy images via deep learning

BRYCE MANIFOLD, ELENA THOMAS, ANDREW T. FRANCIS, ANDREW H. HILL, AND DAN FU*

Department of Chemistry, University of Washington, Seattle, WA 98195, USA

*danfu@uw.edu

Abstract: Stimulated Raman scattering (SRS) microscopy is a label-free quantitative chemical imaging technique that has demonstrated great utility in biomedical imaging applications ranging from real-time stain-free histopathology to live animal imaging. However, similar to many other nonlinear optical imaging techniques, SRS images often suffer from low signal to noise ratio (SNR) due to absorption and scattering of light in tissue as well as the limitation in applicable power to minimize photodamage. We present the use of a deep learning algorithm to significantly improve the SNR of SRS images. Our algorithm is based on a U-Net convolutional neural network (CNN) and significantly outperforms existing denoising algorithms. More importantly, we demonstrate that the trained denoising algorithm is applicable to images acquired at different zoom, imaging power, imaging depth, and imaging geometries that are not included in the training. Our results identify deep learning as a powerful denoising tool for biomedical imaging at large, with potential towards *in vivo* applications, where imaging parameters are often variable and ground-truth images are not available to create a fully supervised learning training set.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Stimulated Raman scattering (SRS) microscopy is a powerful optical imaging technique that uses the intrinsic vibrational contrast of molecules to provide chemical maps of biological cells and tissues. Due to its label-free imaging capability and subcellular spatial resolution, SRS imaging has shown great promise in many biological and biomedical applications such as metabolic studies, drug imaging, tissue diagnosis [1–8]. Recent work has shown that two-color SRS imaging can generate virtual H&E images that are useful for intraoperative cancer detection and margin analysis [9,10]. *In vivo* SRS imaging has also been demonstrated to be useful for drug pharmacokinetics studies [11–13].

Despite recent advances, SRS imaging still faces shortcomings that prevent it from becoming more widely used in biological imaging. Similar to multiphoton fluorescence, SRS imaging uses ultrashort laser pulses to excite weak nonlinear optical transitions. The signal to noise ratio (SNR) decreases rapidly as light is scattered when imaging deeper in a sample. Due to the often weak Raman cross-sections of biomolecules, SRS images often are noisy and of low quality when imaging deep and/or at high speed. For example, SRS signal is often unacceptably noisy when imaging tissue at depths below 90 μm even for high abundance biomolecules such as proteins and lipids [14,15]. Additionally, low signal conditions may be inevitable when considering the limitation of laser power for *in vivo* imaging to avoid tissue damage [16,17]. *In vivo* experiments with SRS imaging are also often inherently signal-limited where epi detection is typically required. Epi imaging acquires back-scattered light from the sample which yields significantly weaker signal in comparison to transmission mode detection. These challenges (depth, laser power, and detection scheme) are common in biological imaging and often result in the acquisition of low SNR images. While in some applications this can be mitigated by increasing imaging times, *in vivo* applications require rapid image collection to both observe physiologically relevant processes and avoid imaging

artifacts due to sample motion. As a result, many *in vivo* studies turn to denoising algorithms to increase their image SNR [18–20].

While standard denoising algorithms can be used to improve image quality, they typically require either *a priori* knowledge about the interfering noise or multiple images of the same features to enable averaging which often introduces undesirable consequences such as a decrease in the effective spatial resolution of the image [21–23]. Recently, deep learning via CNNs has shown significant promise as a denoising tool [24–26]. These CNN-based algorithms have been used to denoise images with inherent compression corruption or induced Gaussian noise, often even performing well in blind denoising tests. However, the fully-connected architecture of the most common CNNs for denoising involve significant training times and require large training samples to be effective. Moreover, these deep learning denoising algorithms are based on RGB images with relatively narrowband noise (noise centered around a small frequency range) [27]. To our knowledge, this is the first report of using a CNN to denoise nonlinear optical images, particularly SRS images.

Here, we report the use of a U-Net architecture CNN to denoise SRS images in low signal situations. Previous work with this U-Net architecture has been able to create algorithms that predict label-free fluorescence images from brightfield microscopy images with high fidelity while requiring relatively few training images [28]. The use of a CNN presents an elegant way of tailoring a specialized denoising algorithm to significantly improve the quality of SRS images in situations where low signal is unavoidable. In this work, we train a deep learning algorithm with corresponding SRS images taken at low and high laser power (i.e. images with low and high signal to noise ratios) then use the trained algorithm to denoise new images with similarly low SNR. Our method significantly outperforms other denoising methods with respect to many common noise and image fidelity metrics. Moreover, we find that the trained algorithm is applicable to images acquired at differing fields of view, imaging powers, imaging depths, and even experimental geometries (epi versus transmission) than the images used to train the algorithm. Lastly, we note that while the denoising algorithm is demonstrated for SRS imaging, it should be equally applicable to other nonlinear optical imaging techniques, providing a generalizable method to improve tissue imaging quality. Our findings demonstrate the power of CNN-based deep learning as a denoising technique and provide an avenue to significantly improve the quality of biological images acquired in a wide variety of low SNR conditions.

2. Methods

2.1 Sample preparation

HeLa cells were cultured in Dulbecco's modified eagle medium with 10% fetal bovine serum at 37 °C with 5% CO₂ atmosphere. Cells were seeded on coverslips 24 hours prior to being fixed using 1% paraformaldehyde.

Murine brain tissue was harvested from recently sacrificed animals provided by UW Animal Use Training Services (AUTS) according to IACUC protocol 3388–03. After excision, thin (~200 μm) sections of tissue were collected and mounted on glass microscope slides.

2.2 SRS imaging

SRS images were acquired using a homebuilt SRS microscope as described previously [29–31] and as shown in Fig. 1. The laser used is a femtosecond dual-output Spectra-Physics Insight DeepSee + which emits a tunable beam (680–1300 nm) and fixed beam 1040 nm pulse trains at a synchronized 80MHz repetition rate. For HeLa cell imaging, a spectral focusing approach, described elsewhere [32], is adopted. The 800 nm pump pulse is chirped using high density glass, while the Stokes pulse (centered at 1040 nm) is stretched using a grating-based pulse stretcher [33]. HeLa cell images were collected at 2913 cm⁻¹ using a spectral resolution

of 15 cm^{-1} . In the two-color mouse brain images two 1040 nm pulse trains, modulated 90° out of phase with each other, are used to enable simultaneous two-color acquisition as shown previously [34,30]. Murine brain images were collected at 2913 cm^{-1} and 2994 cm^{-1} with a spectral resolution of 45 cm^{-1} . The microscope used is a Nikon Eclipse FN1 equipped with a $40\times 1.15\text{ NA}$ water immersion objective for HeLa images and a $25\times 1.05\text{ NA}$ water immersion objective for the tissue images. In the HeLa cell imaging, the pump beam (800 nm) power was held constant at 20 mW at focus, and images were taken using 1 mW and 20 mW of Stokes power for each field of view. In the mouse brain imaging, the pump beam was held at 20 mW and images were taken at either 1 mW or 15 mW each for both Stokes beams for all fields of view. All images collected were 512×512 pixels.

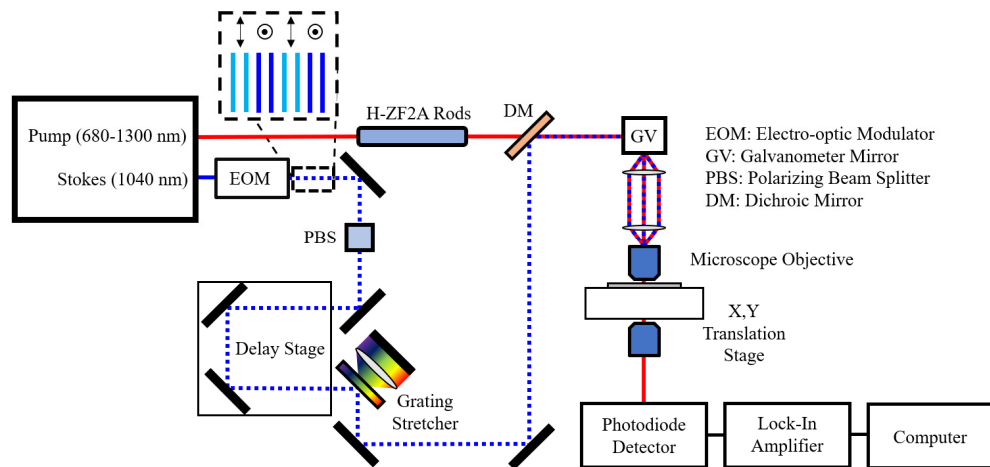


Fig. 1. SRS microscope setup with spectral focusing. Epi detection used the same setup with the addition of a polarizing beam splitter prior to the objective to collect back scattered light and send it to the photodiode detector.

2.3 Deep learning training and denoising

While an in-depth explanation of the functions of a CNN is beyond the scope of this manuscript, we find it prudent to provide some context as to how we believe the CNN we employ operates. The U-Net CNN creates a series of filters that images get passed through as they are successively broken down into lower resolution components. The filters used to break down the initial image are then subjected to a long training process minimize the mean square error (MSE) between the high-power, low noise truth images and the prediction based on the low-power, high noise images. The expectation is that some filters learn to address “macro” effects (optical aberration, nonuniform illumination, object shape and size, etc.) while others address “micro” effects (Poisson noise, pixel-to-pixel variations, fine structural features, etc.). Those filters accounting for macro effects likely reside higher in the architecture, where the image resolution is still comparatively high and the large portions of the image are considered collectively. Macro effects are sample specific, and much of the structural information utilized in the U-Net CNN training process is likely dealt with in these high level filters. Filters accounting for micro effects likely lie lower in the architecture, where image resolution has been significantly reduced and smaller portions of the initial image are considered. Because of the structural specificity learned by the CNN model, best results are only attainable on a model which has been trained on the specific system it is being asked to denoise. For this reason, different models were trained for each of the systems (HeLa cells and murine brain) studied in this work.

The U-Net CNN used in this work was created by Ounkomol et al [28] with small optimizations made for our specific applications. Corresponding low and high SNR images

(i.e. low/high power or epi/transmission modality) were used without any pre-processing to train a denoising algorithm over 50,000 epochs. All deep learning algorithms were supplied 40 fields of view for training with a randomized 10/30 test/train split. The CNN used in this work utilizes a four-layer network. Each layer consists of two 3x3 kernel convolutions followed by batch normalization and a ReLU activation function, then 2-pixel convolutions followed by batch normalization. Our CNN employs a learning rate of 0.001 with an Adam optimizer, momentum values of 0.5 and 0.999, and a batch size of 20 images.

In the case of the two-color mouse brain images, lipid and protein images were fed simultaneously with additional fields of view withheld for further validation. All images shown here and those used for the relevant peak signal to noise ratio (PSNR), root mean squared error (RMSE), and correlation coefficient (CC) comparisons were not part of the training of the deep learning algorithm. All training sessions and predictions were performed on the University of Washington Hyak Mox supercomputer equipped with an Nvidia P100 GPU. Training sessions lasted ~7 hours depending on training batch and buffer sizes. Utilization of the trained algorithm to denoise batches of images took on average 10 seconds depending on the size of the batch.

3. Results

3.1 Denoising low and high power SRS images of fixed HeLa cells

We first denoised images of HeLa cells acquired at 2920 cm^{-1} with low optical power. Because the SNR is linearly proportional to the Stokes power in SRS imaging [35], images acquired at 1 mW (Fig. 2(A)) display 20-fold lower SNR than those acquired at 20 mW (Fig. 2(D)). Low power images appear grainy with low cellular contrast relative to the background and lacking the expected distinct features such as lipid droplets and nuclei as seen in the corresponding high power SRS image. Pairs of corresponding low and high SNR images similar to those shown Fig. 2(A) and 2(D) provided the signal and truth images for the deep learning algorithm training. The resulting algorithm takes an input low SNR image (Fig. 2(A)) and returns a predicted high SNR image such as the image shown in Fig. 2(C). For comparison, the image shown in Fig. 2(A) is also denoised with variance stabilization transform (VST) Poisson denoising [36], the result of which is shown in Fig. 2(B). To better visualize the ability of the denoising techniques to recover cellular features and background contrast, pixel value plots along the same line region in all images are shown in Fig. 2(E). An image of the same field of view denoised using PURE-LET is provided in the appendix (Fig. 8). In the low SNR plot (orange), the variation in pixel value along the image dominates the cellular features more clearly seen in the high SNR line plot (red). While these spatial features are partially recovered and the variation from the noise is suppressed in the VST denoised plot (magenta), the sharp features (cell edges, lipid droplets, nucleoli, etc.) are not well recovered and significantly blurred. The deep learning algorithm, however, demonstrates significant denoising of the low SNR image with near perfect separation of the cells from the background and significant recovery of cellular features such as lipid droplets and nuclei.

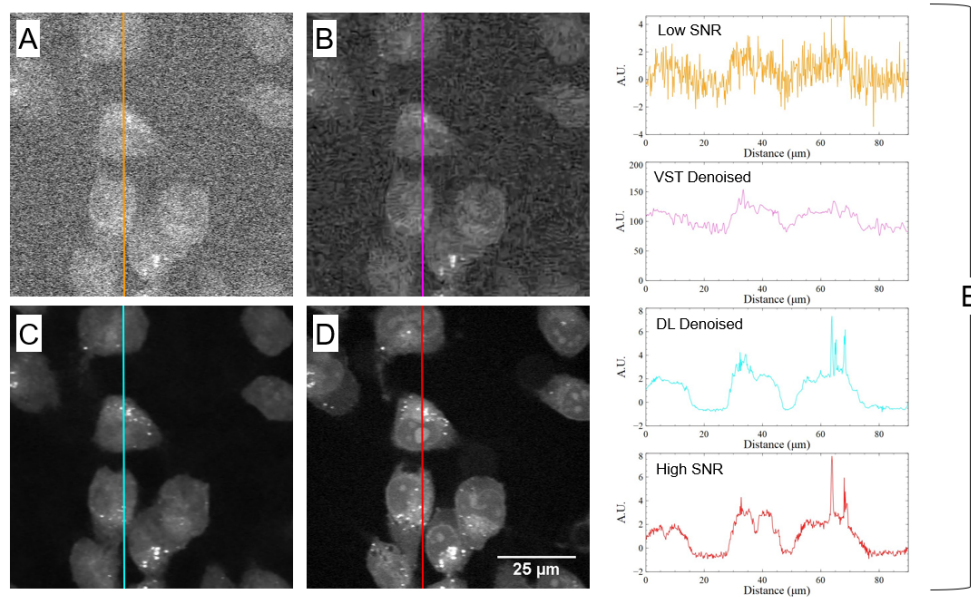


Fig. 2. (A-D) SRS images at 2920 cm^{-1} of fixed HeLa cells where (A) was imaged using 1mW Stokes and 20 mW pump (low power) then subsequently denoised using VST (B) and the trained deep learning algorithm (C); (D) shows the same field of view as (A-C) imaged at 20 mW Stokes and pump (high power). Corresponding pixel value plots along the shown line are demonstrated in (E).

To further quantify the denoising capability of the deep learning algorithm in comparison to other denoising methods, PSNR, RMSE, and CC values were calculated with respect to the high power images (equations shown in appendix). PSNR is a metric that expresses a logarithmic measure of image quality with respect to a truth image. High PSNR indicates higher image fidelity. RMSE expresses the accuracy of the denoising method with respect to a truth image. A low RMSE indicates an accurate denoising method. CC is the Pearson correlation coefficient that expresses colocalization of features in the test and truth images as a number between -1 and 1 . A CC of -1 , 0 , or 1 would indicate perfect anti-correlation, no correlation, or perfect correlation respectively. These values were calculated in ImageJ using previously written plugins [37] with the withheld test images to avoid concerns of overfitting. As shown in Table 1, the deep learning denoising significantly outperforms other denoising algorithms optimized for removing Poisson-shaped noise (VST and PURE-LET [36,38]). VST denoising slightly decreases PSNR and increases RMSE in comparison to the original input image likely due to the extremely low starting PSNR and significant blurring of spatial features during the denoising process.

Table 1. PSNR, RMSE, and CC^a Values of Low Power HeLa Images and Corresponding Denoising Methods

Images	PSNR (dB)	RMSE	CC
Low Power	9.99	0.279	0.34
VST Denoised	9.78	0.289	0.59
PURE-LET Denoised	13.53	0.186	0.70
Deep Learning Denoised	28.87	0.049	0.94

^aCalculated with respect to corresponding high power images after normalizing pixel values 0 to 1. Values shown are averages from 10 images.

Overall, our data suggests that given the correct training (even on a relatively small training set of 30 images), the U-Net CNN deep learning algorithm demonstrates strong capability for denoising images taken using low powers.

3.2 Denoising low and high power two-color SRS images of ex vivo mouse brain

To further assess the utility of deep learning denoising in SRS imaging of tissue, we utilized our algorithm to denoise murine coronal brain tissue section images. While images of fixed HeLa cells have well-defined internal features and a truly signal-free background (parameters which U-Net architectures excel in learning [39]), murine brain tissue exhibits significantly more heterogeneity and no true background. In this case our algorithm was trained with corresponding low and high power two-color SRS images acquired in transmission mode.

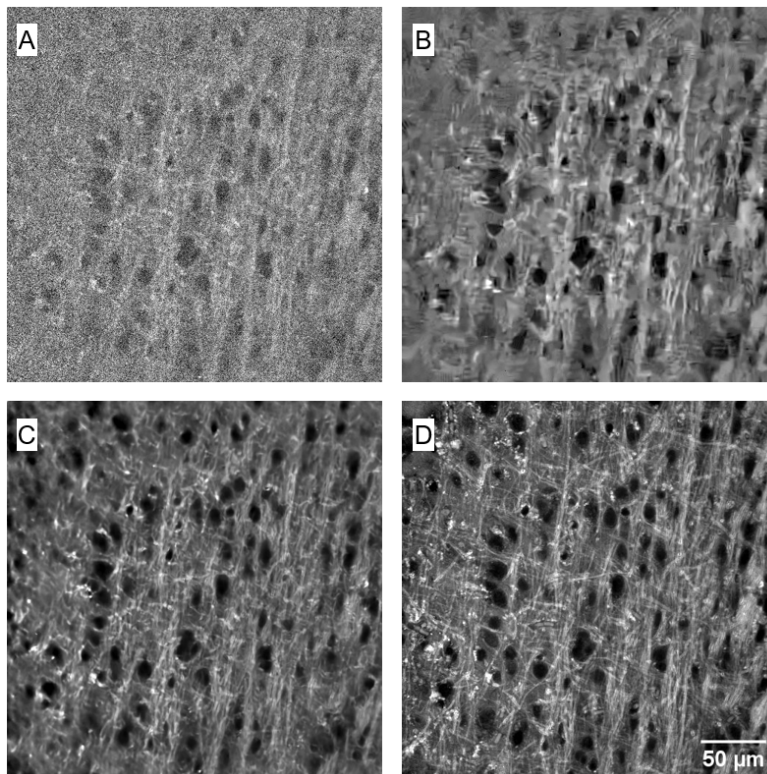


Fig. 3. Coronal mouse brain SRS images acquired at 2990 cm^{-1} . (A) Low power image acquired at 1 mW Stokes and 20 mW pump. (B) The low power image denoised with VST. (C) The low power image denoised with the deep learning algorithm. (D) The high power image acquired at 20 mW Stokes and 20 mW pump.

The relative denoising capability of the algorithm is first shown in Fig. 3 with only lipid channel (2990 cm^{-1}) of the two-color images. The lipid images shown in grey scale demonstrate the high fidelity of denoising by the deep learning-trained algorithm. Figures 3(A) and 3(D) show the low and high power images for the field of view respectively. Figures 3(B) and 3(C) show the VST and deep learning denoised versions of the low power image respectively. It is clear given the low SNR of the initial image, that VST again significantly blurs spatial features in comparison to the high power image. The deep learning denoised image, however, demonstrates significant denoising without the loss significant loss in spatial resolution, especially in the center of the image. Spatial field of view heterogeneity is common in SRS imaging due to chromatic aberration near the edges of the image. VST denoising fails to recover features around the edges while deep learning does reasonably well.

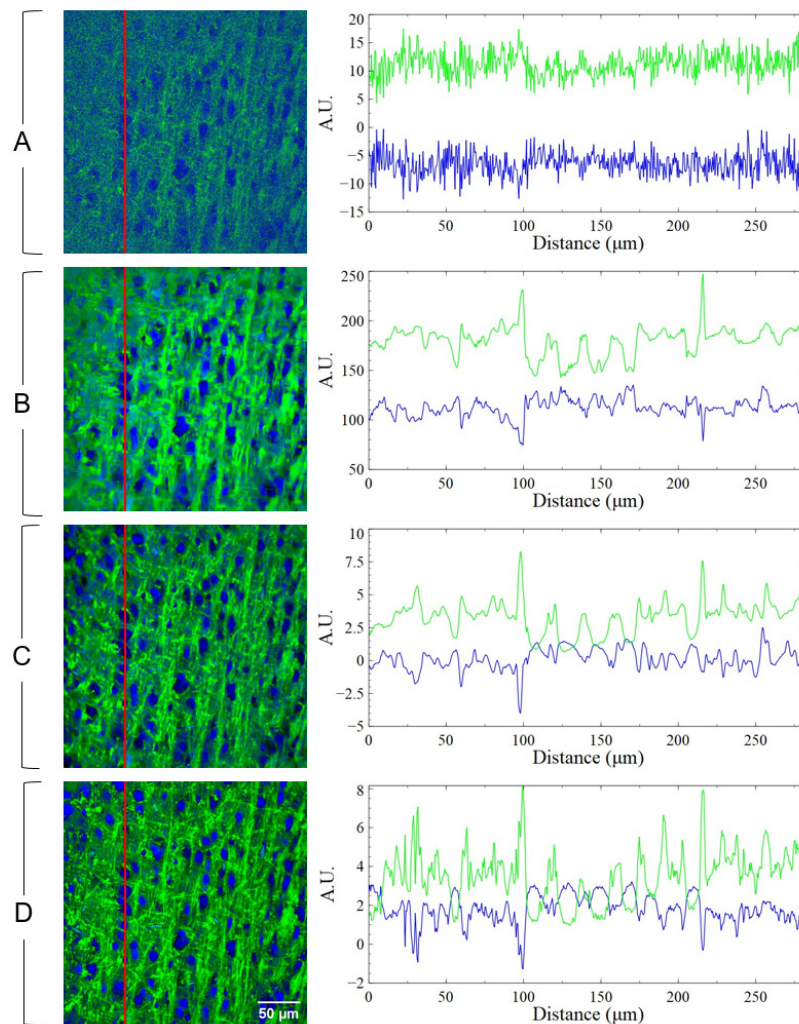


Fig. 4. Two-color (lipids and proteins) composite SRS images of *ex vivo* mouse brain and corresponding pixel value line plots. (A) Low power image acquired at 1 mW Stokes 20 mW pump. (B) The low power image denoised using VST. (C) The low power image denoised with the deep learning algorithm. (D) The high power image acquired at 20 mW Stokes and pump. Pixel value line plots along the red line are shown for each composite image.

The two-color SRS images (where lipids are colored green and proteins are colored blue) at an imaging depth of $\sim 10 \mu\text{m}$ are shown in Fig. 4. In two-color SRS images, significant contrast between the lipid and protein transitions can be used to generate diagnostic maps for pathology applications [6,9,40]. In the low power images (Fig. 4(A)) however, such contrast is absent, making low SNR images inadequate for pathology. While both VST denoising (Fig. 4(B)) and CNN denoising (Fig. 4(C)) recover the nuclei contrast, the latter performed significantly better in terms of imaging fidelity and spatial resolution. The pixel value plots along the shown lines for both the lipids and protein channels are shown next to their respective images. Here it is evident that, despite the noise in the initial low power image and the heterogeneity in tissue features, deep learning significantly recovers many of the sharp features (mostly axons in green and nuclei in blue) visible in the high power image. While VST does remove a significant portion of the noise, features are blurred, especially around the edges of the image (as seen in Fig. 2 and Fig. 3). Further, the relatively narrow axons evident

in the line plots are recovered with high fidelity (Fig. 4(C)) indicating deep learning denoising does not sacrifice spatial resolution as other denoising techniques do.

Analysis of PSNR, RMSE, and CC values in these two-color images (shown in Table 2) reveals a similar trend to that seen in the HeLa cells. That is, deep learning significantly outperforms other Poisson denoising methods across all metrics for these low power images. While the CC value doesn't approach unity as in the HeLa images, this is likely due to the more heterogeneous background in the tissue compared with the fixed HeLa cell images.

Table 2. PSNR and RMSE^a Values of Low Power and Denoised Two-Color SRS Mouse Brain Images

Images	PSNR (dB)	RMSE	CC
Low Power	18.04	0.133	0.59
VST Denoised	17.18	0.141	0.73
PURE-LET Denoised	19.75	0.112	0.74
Deep Learning Denoised	25.24	0.066	0.82

^aCalculated with respect to corresponding high power images after normalizing pixel values 0 to 1. Values shown are averages from 10 images including lipid and protein channels.

This data demonstrates that U-Net based deep learning can create a powerful denoising algorithm with relatively small training sets not only for low power SRS imaging of fixed cells, but also of heterogenous tissue samples. This suggests the possibility of using deep learning to further enhance the current capabilities of SRS imaging.

3.3 Denoising deep SRS images of ex vivo mouse brain

One major limitation of deep learning based denoising approaches is blind denoising of images markedly different from those in the training set. This would pose a potential limitation to denoising SRS images deep in tissue as adequate supervised-learning training sets cannot be created due to the inherently low SNR. However, because SRS images at any depth share the same noise features, we hypothesize that we can apply the algorithm trained at shallower imaging depths to images deeper into tissue. We tested this approach by using the algorithm trained in the previous section to SRS images of the same tissue at depths up to 175 μm , a depth that has not been reached before in previous reports of native SRS imaging. In this validation, two-color images were taken at high power (15/20 mW Stokes/pump for both channels) at depths of up to 175 μm into the mouse brain tissue. They were then denoised using the previously shown VST denoising and CANDLE denoising [41]. Images at a depth of 175 μm and denoising of the acquired SRS image are shown in Fig. 5. Here, there is no truth high SNR image with which to compare. The SRS image taken at high power is shown in Fig. 5(A) and denoised via VST, CANDLE, and the trained deep learning algorithm are shown in Figs. 5(B)-5(D) respectively. The corresponding pixel value line plots are shown for each image. Figure 5(A) demonstrates the loss in SNR as images are acquired deeper into tissue, even when higher powers are used. From the images and line plots it is clear VST and CANDLE do not fully remove the noise inherent to SRS images at this depth. The deep learning denoising significantly removes the noise, resulting images and pixel plots similar to the denoised low power images at shallower depths (Figs. 3 and 4). Specifically, the deep learning denoising recovers the expected axons in the lipid channel and better resolves nuclei from the background in the protein channel without significant blurring of any features.

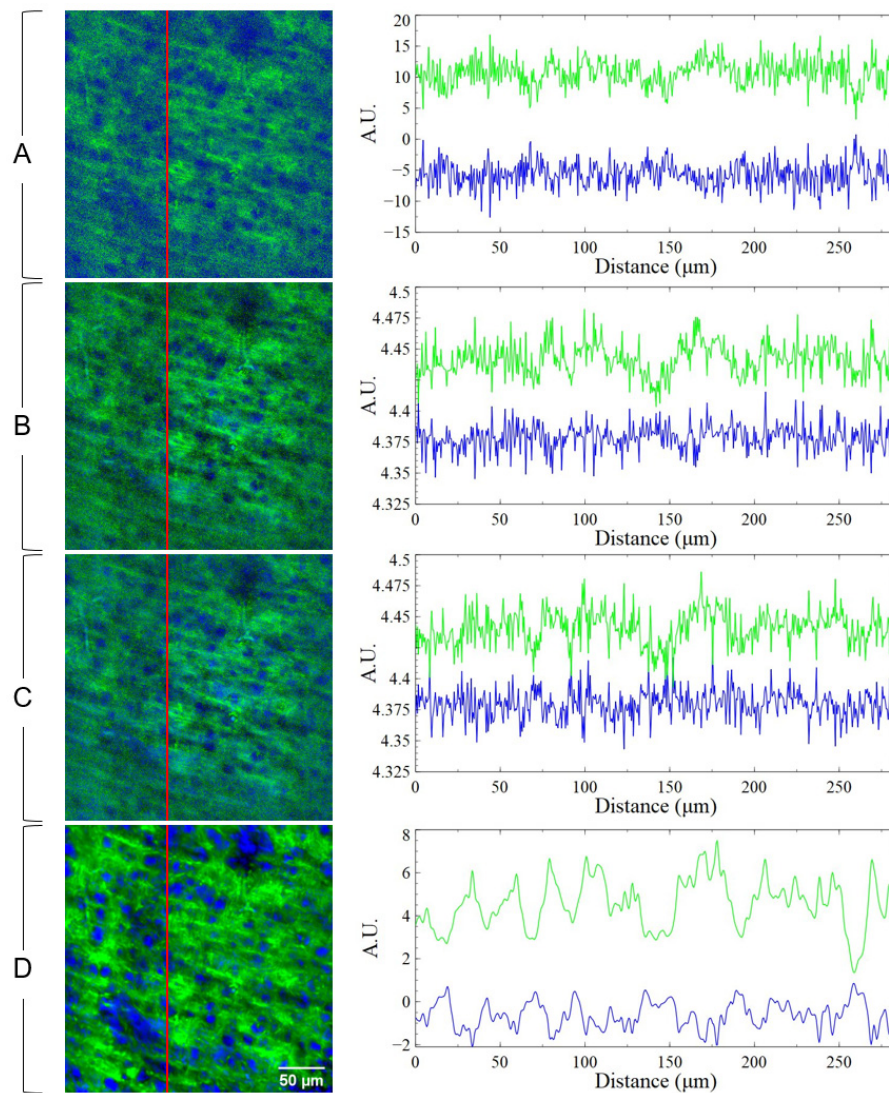


Fig. 5. (A-D) Two-color (lipids-green, proteins-blue) SRS images of coronal mouse brain slice at a depth of 175 μm. (A) shows the field of view taken at 15 mW Stokes 20 mW pump (high power) and (B-D) show denoising of (A) via VST (B), CANDLE (C), and the trained deep learning algorithm (D). Corresponding pixel plots along the lines shown next to their respective images.

The recovery of signal in both the lipid and protein channels demonstrates that the training of the deep learning algorithm at low and high powers in shallower tissue imaging can improve images acquired at much deeper depth. While the laser powers used here are the same as in the high power images shown previously, the images collected have inherently low SNR due to power lost to tissue absorption and scattering. In SRS imaging where shot noise is expected to be the limiting factor in low signal regimes, it would follow that a deep learning algorithm trained to recover signal among this shot-noise would effectively improve imaging depths. The validation shown in Fig. 5 suggests that the algorithm is robust in denoising these two-color SRS images, despite not having explicitly trained on images this deep into the brain. Furthermore, it is unnecessary to acquire multiple images of a given field of view to create an average denoised image. CANDLE, for example, requires multiple

images over which to learn the average noise distribution. In this case one need only provide a reasonable training set from which the algorithm may learn, then any single image may be denoised using the trained algorithm.

To further examine the generalizability of the algorithm trained in Fig. 4, the algorithm was also validated with a low power image taken with a $142 \times 142 \mu\text{m}$ field of view, compared to the $285 \times 285 \mu\text{m}$ fields of view used in the training set. As shown in Fig. 6(A)-6(C), the algorithm remains effective in denoising the low power image even at a zoom different from the images utilized for the training set. The axons and nuclei are reliably recovered without any significant augmentation of shape or size. This further indicates that while the model likely uses structural information in its predictions it does not strictly impose the object size distributions native to its training set. This suggests that a well-trained algorithm would be widely applicable for low SNR situations of a given sample.

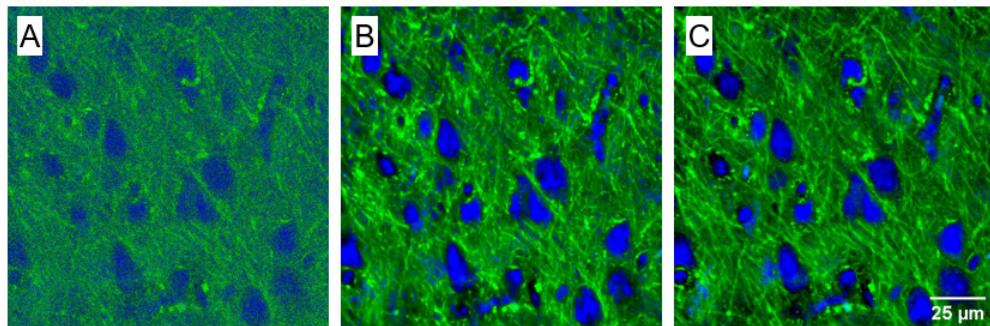


Fig. 6. Two-color (lipids-green, proteins-blue) SRS images of coronal mouse-brain slice. Images (A-C) show an area imaged at 2x zoom where (A) was imaged at low power, (B) was the low power image denoised using the same deep learning algorithm from Figs. 3 & 4, and (C) was imaged at high power.

3.4 Denoising epi two-color SRS images of mouse brain with transmission images trained algorithm

Finally, we examine the ability of deep learning to create an algorithm that can improve the quality of epi-SRS images to the level of transmission SRS images. This capability is particularly important for *in vivo* imaging, where epi-imaging is required due to the opacity and thickness of many samples. Less light is recovered and directed towards the detector in epi-SRS imaging with respect to the transmissive experiment. This results in a significant loss in SNR. We hypothesize that we can apply training algorithm obtained from transmission images to improve epi-SRS imaging and reduce the image quality discrepancy between the two experiments.

Figure 7 depicts the application of our denoising algorithm, trained with transmissive SRS images to the denoising of images collected using an epi-SRS geometry. The images shown in Figs. 7(A) and 7(C) show high power two-color images of murine brain at a depth of $15 \mu\text{m}$ in epi and transmission mode respectively. Figure 7(B) shows the epi image denoised by the deep learning algorithm.

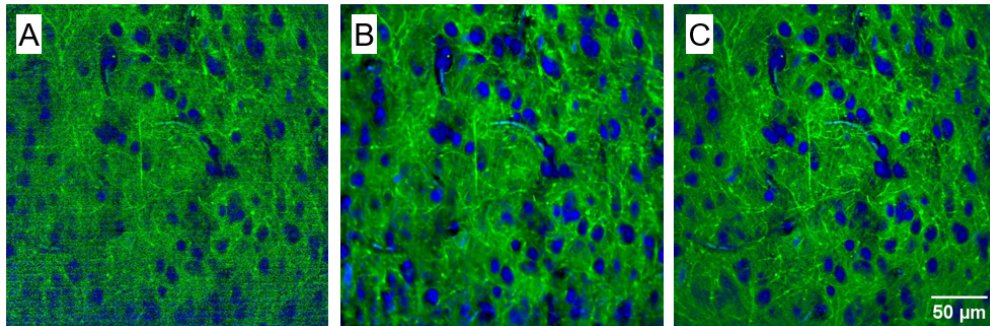


Fig. 7. Images (A-C) show an area of coronal mouse brain imaged in epi mode (A), the epi field of view denoised via deep learning algorithm (B), and the same field of view imaged in transmission mode (C).

Figure 7(A) demonstrates how the noise starts to obfuscate the signal features even at high powers near the surface of the tissue in epi-SRS. In applying this algorithm to the epi-SRS images, similar to the demonstrations shown in Figs. 1-6, image quality is significantly improved, approaching the quality of simultaneously acquired transmission-SRS images (Fig. 7(B-C)). This suggests an exciting method of improving *in vivo* epi-SRS imaging, whereby the epi images can be improved and denoised by a pre-trained algorithm using *ex vivo* tissue. This would effectively increase the imaging depth and utility of SRS imaging *in vivo*.

4. Discussion

SRS imaging is a powerful label-free imaging technique that provides chemical information at sub-micron resolution. When using SRS imaging to examine biological systems, however, some limitations become evident. Scattering and absorption in tissue attenuates signal which ultimately limits image quality and imaging depths. Additionally, tissue can also be damaged by high power lasers but lowering laser power conversely lowers the signal. Finally, the requirement of an epi imaging modality in *in vivo* applications inherently reduces the collected SRS signal as backscattered light is weaker than forward propagating light in transmission modality. These limitations stem from the same basis problem: signal strength is limited in the collection of an image resulting in noise dominated images.

While many general denoising algorithms have been developed for removing Poisson shaped shot-noise (mostly in the context of fluorescence imaging), none are aptly suited for the extremely low SNR's observed in the limits of SRS imaging. The general denoising algorithms fail to recover the inherent quantitative information in SRS images and often blur the relevant biological features.

Here we demonstrate the first use of deep learning to denoise and improve the quality of SRS images that outperforms more general denoising methods. Given the appropriate training, the deep learning denoising algorithms demonstrated here spatially recover relevant biological features (e.g. lipid droplets, axons, nuclei, etc.) without blurring or overfitting of features. Additionally, the deep learning algorithms appear to recover appropriate pixel values for truth images indicating the potential for recovery of quantitative information (pixel value plots from Figs. 2, 4, and 5). The main limitation of using a deep learning algorithm is the necessitation of acquiring an appropriate training set and the inherent trade off in generalizability of denoising in comparison to other denoising algorithms such as the here shown VST, PURE-LET, and CANDLE methods. For example, denoising the HeLa images with the mouse brain-trained algorithm or vice-versa exhibits worse performance than the appropriately trained algorithm for the given system. This trade-off, however, does not detract from the overall performance of the deep learning trained algorithm when used appropriately, especially considering the relatively small training sets used in these experiments (30 images for training). It is also worth considering the relative generalizability of the trained algorithm

of the system. For example, we demonstrated that training a single 30 image data set is applicable to images acquired at different zoom, imaging depth, imaging power, and even imaging geometry. This demonstrates that the U-Net based algorithm created here should be treated as a specialized tool for improving SRS imaging. Additionally, while all results shown here are acquired with *ex vivo* samples, there is strong indication that *in vivo* images could be similarly improved given appropriate algorithm training (such as that shown in Fig. 7).

Ultimately, deep learning is valuable in augmenting the capabilities of SRS imaging in biological systems. Specifically, deep learning can improve the depths at which native biological information may be recovered, *in vivo* tissue imaging, and imaging of biomolecules at low abundance. The generalizability of this deep-learning based denoising approach may be improved in future work aimed at training a CNN on the noise profile in low-power SRS images directly. Additionally, utilizing a structural similarity (SSIM) loss function rather than an MSE loss function may provide more robust results [42]. That will be the subject of future studies. We expect future deep learning work in SRS imaging to expound on these improvements towards improving the utility of SRS imaging.

Appendix 1

Eq. (1) shows how PSNR is calculated where r and t are the reference (truth) and tested (denoised) image respectively, (x,y) is a given pixel coordinate in an $n_x \times n_y$ image, and R is the maximum value of the tested image. Images that contain more noise will have lower PSNR values.

$$PSNR = 10 \log_{10} \left(\frac{R^2}{\sum_{r,t} [r(x,y) - t(x,y)]^2 / n_x * n_y} \right) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n_x n_y} \cdot \sum_0^{n_x-1} \sum_0^{n_y-1} [r(x,y) - t(x,y)]^2} \quad (2)$$

Eq. (2) shows how RMSE is calculated. The RMSE is a value that expresses the accuracy of a given denoising method based on a calculation of error at each pixel coordinate between the two images. RMSE values closer to 0 indicate a more accurate denoising. Note that PSNR implicitly utilizes the RMSE squared or mean square error.

$$CC = \frac{\sum_0^{n_x-1} \sum_0^{n_y-1} [r(x,y) - \bar{r}][t(x,y) - \bar{t}]}{\sqrt{\sum_0^{n_x-1} \sum_0^{n_y-1} [r(x,y) - \bar{r}]^2} \sqrt{\sum_0^{n_x-1} \sum_0^{n_y-1} [t(x,y) - \bar{t}]^2}} \quad (3)$$

Equation (3) shows how CC (also known as the Pearson correlation coefficient) is calculated where \bar{r} and \bar{t} are the average pixel value for the reference (truth) and test (denoised) image respectively. The CC is a measure of the covariance between the reference and test image divided by the product of the standard deviations of the respective images. Here, if features are localized with one another between two images the CC trends towards 1. Noise in the test image results in CC values closer to 0.

Appendix 2

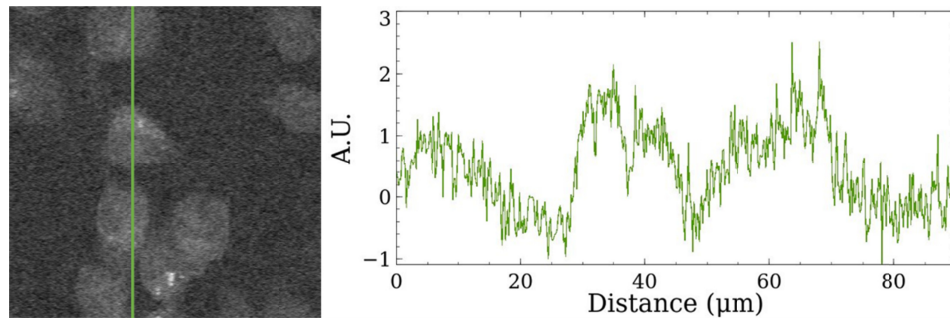


Fig. 8. Image of HeLa cells shown in Fig. 2 denoised using PURE-LET.

Funding

Arnold and Mabel Beckman Young Investigator Award; UW Start-up Fund.

Acknowledgments

We thank Drs. Greg Johnson and Chek Ounkomol for the development and guidance on the U-Net architecture used. We thank Eleanor Munger for kindly culturing the HeLa cells and preparing the samples for imaging.

This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. J.-X. Cheng and X. S. Xie, "Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine," *Science* **350**(6264), aaa8870 (2015).
2. D. Fu, "Quantitative chemical imaging with stimulated Raman scattering microscopy," *Curr. Opin. Chem. Biol.* **39**, 24–31 (2017).
3. L. Shi, C. Zheng, Y. Shen, Z. Chen, E. S. Silveira, L. Zhang, M. Wei, C. Liu, C. de Sena-Tomas, K. Targoff, and W. Min, "Optical imaging of metabolic dynamics in animals," *Nat. Commun.* **9**(1), 2995 (2018).
4. D. Fu, J. Zhou, W. S. Zhu, P. W. Manley, Y. K. Wang, T. Hood, A. Wylie, and X. S. Xie, "Imaging the intracellular Distribution of Tyrosine Kinase Inhibitors in Living Cells with Quantitative Hyperspectral Stimulated Raman Scattering," *Nat. Chem.* **6**(7), 614–622 (2014).
5. B. Figueroa, T. Nguyen, S. Sotthivirat, W. Xu, T. Rhodes, M. S. Lamm, R. L. Smith, C. T. John, Y. Su, and D. Fu, "Detecting and Quantifying Microscale Chemical Reactions in Pharmaceutical Tablets by Stimulated Raman Scattering Microscopy," *Anal. Chem.* **91**(10), 6894–6901 (2019).
6. D. A. Orringer, B. Pandian, Y. S. Niknafs, T. C. Hollon, J. Boyle, S. Lewis, M. Garrard, S. L. Hervey-Jumper, H. J. L. Garton, C. O. Maher, J. A. Heth, O. Sagher, D. A. Wilkinson, M. Snuderl, S. Venneti, S. H. Ramkissoon, K. A. McFadden, A. Fisher-Hubbard, A. P. Lieberman, T. D. Johnson, X. S. Xie, J. K. Trautman, C. W. Freudiger, and S. Camelo-Piragua, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," *Nature Biomedical Engineering* **1**(2), 27 (2017).
7. F.-K. Lu, S. Basu, V. Igras, M. P. Hoang, M. Ji, D. Fu, G. R. Holtom, V. A. Neel, C. W. Freudiger, D. E. Fisher, and X. S. Xie, "Label-free DNA imaging in vivo with stimulated Raman scattering microscopy," *Proc. Natl. Acad. Sci. U.S.A.* **112**(37), 11624–11629 (2015).
8. B. Figueroa, W. Fu, T. Nguyen, K. Shin, B. Manifold, F. Wise, and D. Fu, "Broadband hyperspectral stimulated Raman scattering microscopy with a parabolic fiber amplifier source," *Biomed. Opt. Express* **9**(12), 6116–6131 (2018).
9. M. Ji, S. Lewis, S. Camelo-Piragua, S. H. Ramkissoon, M. Snuderl, S. Venneti, A. Fisher-Hubbard, M. Garrard, D. Fu, A. C. Wang, J. A. Heth, C. O. Maher, N. Sanai, T. D. Johnson, C. W. Freudiger, O. Sagher, X. S. Xie, and D. A. Orringer, "Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy," *Sci. Transl. Med.* **7**(309), 309ra163 (2015).

10. F.-K. Lu, D. Calligaris, O. I. Olubiyi, I. Norton, W. Yang, S. Santagata, X. S. Xie, A. J. Golby, and N. Y. R. Agar, "Label-Free Neurosurgical Pathology with Stimulated Raman Imaging," *Cancer Res.* **76**(12), 3451–3462 (2016).
11. B. G. Saar, C. W. Freudiger, J. Reichman, C. M. Stanley, G. R. Holtom, and X. S. Xie, "Video-Rate Molecular Imaging In Vivo with Stimulated Raman Scattering," *Science* **330**(6009), 1368–1370 (2010).
12. N. L. Garrett, B. Singh, A. Jones, and J. Moger, "Imaging microscopic distribution of antifungal agents in dandruff treatments with stimulated Raman scattering microscopy," *J. Biomed. Opt.* **22**(6), 066003 (2017).
13. W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton, and A. N. Hulme, "Stimulated Raman scattering microscopy: an emerging tool for drug discovery," *Chem. Soc. Rev.* **45**(8), 2075–2089 (2016).
14. M. Wei, L. Shi, Y. Shen, Z. Zhao, A. Guzman, L. J. Kaufman, L. Wei, and W. Min, "Volumetric chemical imaging by clearing-enhanced stimulated Raman scattering microscopy," *Proc. Natl. Acad. Sci. U.S.A.* **116**(14), 6608–6617 (2019).
15. X. Zhang, M. B. J. Roeffaers, S. Basu, J. R. Daniele, D. Fu, C. W. Freudiger, G. R. Holtom, and X. S. Xie, "Label-free live-cell imaging of nucleic acids using stimulated Raman scattering microscopy," *ChemPhysChem* **13**(4), 1054–1059 (2012).
16. M. Nourhashemi, M. Mahmoudzadeh, and F. Wallois, "Thermal impact of near-infrared laser in advanced noninvasive optical brain imaging," *Neurophotonics* **3**(1), 015001 (2016).
17. M. L. Denton, M. S. Foltz, L. E. Estlack, D. J. Stolarski, G. D. Noojin, R. J. Thomas, D. Eikum, and B. A. Rockwell, "Damage Thresholds for Exposure to NIR and Blue Lasers in an In Vitro RPE Cell System," *Invest. Ophthalmol. Vis. Sci.* **47**(7), 3065–3073 (2006).
18. S. Adabi, S. Ghavami, M. Fatemi, and A. Alizad, "Non-Local Based Denoising Framework for In Vivo Contrast-Free Ultrasound Microvessel Imaging," *Sensors (Basel)* **19**(2), E245 (2019).
19. S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," *J. Biomed. Opt.* **14**(1), 014031 (2009).
20. E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. S. Peterka, R. Yuste, and L. Paninski, "Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data," *Neuron* **89**(2), 285–299 (2016).
21. G. Gilboa, Y. Y. Zeevi, and N. Sochen, "Signal and image enhancement by a generalized forward-and-backward adaptive diffusion process," in *2000 10th European Signal Processing Conference* (2000), pp. 1–4.
22. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
23. A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)* (2005), **2**, pp. 60–65 vol. 2.
24. B.-J. Zou, Y.-D. Guo, Q. He, P.-B. Ouyang, K. Liu, and Z.-L. Chen, "3D Filtering by Block Matching and Convolutional Neural Network for Image Denoising," *J. Comput. Sci. Technol.* **33**(4), 838–848 (2018).
25. S. Bako, T. Vogels, B. McWilliams, M. Meyer, J. K. Nová, A. Harvill, P. Sen, T. Deroose, and F. Rousselle, "Kernel-predicting Convolutional Networks for Denoising Monte Carlo Renderings," *ACM Trans. Graph.* **36**(4), 97 (2017).
26. S. K. Choudhury, P. K. Sa, R. P. Padhy, and B. Majhi, "A denoising inspired deblurring framework for regularized image restoration," in *2016 IEEE Annual India Conference (INDICON)* (2016), pp. 1–6.
27. D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach," *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18* (2017).
28. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Methods* **15**(11), 917–920 (2018).
29. B. Figueroa, Y. Chen, K. Berry, A. Francis, and D. Fu, "Label-Free Chemical Imaging of Latent Fingerprints with Stimulated Raman Scattering Microscopy," *Anal. Chem.* **89**(8), 4468–4473 (2017).
30. A. Francis, K. Berry, Y. Chen, B. Figueroa, and D. Fu, "Label-free pathology by spectrally sliced femtosecond stimulated Raman scattering (SRS) microscopy," *PLoS One* **12**(5), e0178750 (2017).
31. A. T. Francis, T. T. Nguyen, M. S. Lamm, R. Teller, S. P. Forster, W. Xu, T. Rhodes, R. L. Smith, J. Kuiper, Y. Su, and D. Fu, "In Situ Stimulated Raman Scattering (SRS) Microscopy Study of the Dissolution of Sustained-Release Implant Formulation," *Mol. Pharm.* **15**(12), 5793–5801 (2018).
32. D. Fu, G. Holtom, C. Freudiger, X. Zhang, and X. S. Xie, "Hyperspectral Imaging with Stimulated Raman Scattering by Chirped Femtosecond Lasers," *J. Phys. Chem. B* **117**(16), 4634–4640 (2013).
33. V. Chauhan, P. Bown, J. Cohen, and R. Trebino, "Single-diffraction-grating and grism pulse compressors," *J. Opt. Soc. Am. B* **27**(4), 619–624 (2010).
34. R. He, Y. Xu, L. Zhang, S. Ma, X. Wang, D. Ye, and M. Ji, "Dual-phase stimulated Raman scattering microscopy for real-time two-color imaging," *Optica*, **OPTICA** **4**(1), 44–47 (2017).
35. C. Zhang, D. Zhang, and J.-X. Cheng, "Coherent Raman Scattering Microscopy in Biology and Medicine," *Annu. Rev. Biomed. Eng.* **17**(1), 415–445 (2015).
36. L. Azzari and A. Foi, "Variance stabilization in Poisson image deblurring," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (2017), pp. 728–731.
37. D. Sage and M. Unser, "Teaching image-processing programming in Java," *IEEE Signal Process. Mag.* **20**(6), 43–52 (2003).

38. J. Li, F. Luisier, and T. Blu, "Deconvolution of poissonian images with the PURE-LET approach," in *2016 IEEE International Conference on Image Processing (ICIP)* (2016), pp. 2708–2712.
39. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., Lecture Notes in Computer Science (Springer International Publishing, 2015), pp. 234–241.
40. L. Zhang, Y. Wu, B. Zheng, L. Su, Y. Chen, S. Ma, Q. Hu, X. Zou, L. Yao, Y. Yang, L. Chen, Y. Mao, Y. Chen, and M. Ji, "Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy," *Theranostics* **9**(9), 2541–2554 (2019).
41. P. Coupé, M. Munz, J. V. Manjón, E. S. Ruthazer, and D. L. Collins, "A CANDLE for a deeper in vivo insight," *Med. Image Anal.* **16**(4), 849–864 (2012).
42. H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," *IEEE Trans. Comput. Imaging* **3**(1), 47–57 (2017).